14

MULTIVARIATE STATISTICAL PROCESS CONTROL

YUAN YAO

Department of Chemical Engineering, National Tsing Hua University, Hsinchu, Taiwan

FURONG GAO

Department of Chemical and Biomolecular Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong, China

In the past several decades, tremendous advancements have been made in automatic process control (APC), which is also called *engineering process control (EPC)*. Although the applications of APC reduce a large part of short-term variations in industrial processes, it is still necessary to detect the abnormal operating conditions and to diagnose the corresponding causes so as to improve the process performance in a long term. Industrial statistics show that even though major catastrophes and disasters are infrequent, minor accidents are very common, resulting in many occupational injuries, and illnesses, and leading to large economic losses every day.[1] Therefore, process monitoring and fault diagnosis are essential for ensuring process safety and product quality consistency.

14.1 STATISTICAL PROCESS CONTROL

In process monitoring, a model describing nominal operating conditions based on which an in-control region (normal operating region) can be defined is usually required. Comparing the real status of a process with the model, process abnormalities can be detected as a significant departure from the in-control region. Generally, a process can be modeled based on first principle, expert knowledge, or historical data. Owing to process complexities, it is difficult to build an accurate first-principle model, while expert knowledge is often not sufficient for experiential modeling. At the same time, with rapid developments of electrical sensors and computer technology, it is easy to get large amounts of process measurements from previous operations. Knowledge of process operating status and product qualities is inferred from these process data. The data-based process models with which the on-line monitoring and fault diagnosis are performed can be easily built. Therefore, statistical process control (SPC), which extracts process characteristics from large samples and defines in-control region through hypothesis testing method, attracts increasing research and application interests.

Control charts are the most widely utilized tools in SPC. In traditional univariate SPC, the applications of Shewhart chart,[2] cumulative sum (CUSUM) chart,[3,4] exponentially-weighted moving average (EWMA) chart,[5] and so on, have been well accepted by many industries, including the polymer processing industry.[6] However, a univariate SPC chart can only reflect the status of a single process variable, while the data from real industrial processes are always high dimensional. Therefore, the monitoring task becomes troublesome for technicians as they have to observe a large number of control charts at the same time. Moreover, when process variables are not independent of each other, none of them adequately define process normality by itself. Under such a situation, univariate SPC does not have the ability to extract variable correlation information, which increases the possibility of miss-alarms because variable correlations are important process characteristics and indicate operation normality. This point can be demonstrated using an

Computer Modeling for Injection Molding: Simulation, Optimization, and Control, First Edition. Edited by Huamin Zhou.

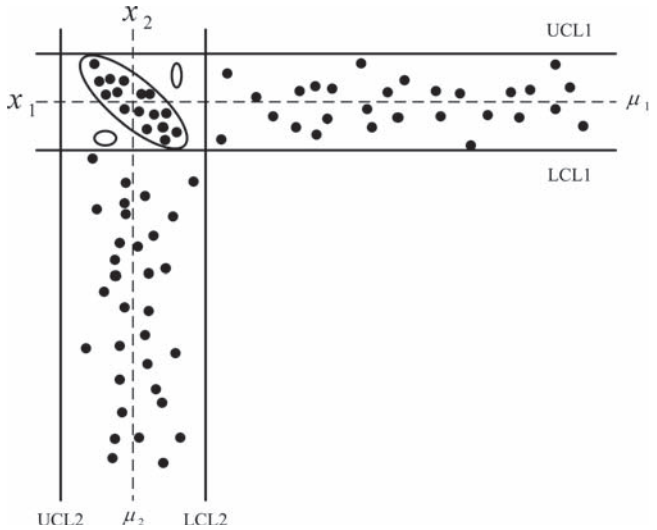

FIGURE 14.1 Monitoring of two correlated variables.

illustrative example shown in Figure 14.1. Suppose there are two variables to be monitored, x_1 and x_2. Because of the existence of variable correlations, the normal operating region of this process should be defined by the combined distribution of x_1 and x_2, which is located in an elliptical region as shown in the figure. The same observations are also plotted as individual Shewhart charts on x_1 and x_2. μ_1 and μ_2 are mean values of x_1 and x_2, while UCL1, LCL1, UCL2, and LCL2 are the control limits for the corresponding Shewhart charts. The small circle in the plot is an outlier obviously outside the normal operating region, but cannot be detected by any of the individual Shewhart charts. Such an example reveals the necessity of applying multivariate statistical process control (MSPC) methods.

14.2 MULTIVARIATE STATISTICAL PROCESS CONTROL

14.2.1 Principal Component Analysis

To deal with high dimensional and highly correlated process data, principal component analysis (PCA)[7,8] has become a key technique in MSPC. Assuming that the process data space is observed with measurement noises effectively in low dimension, PCA has the ability to reduce data dimensions without losing useful information. Many related research works focused on different process characteristics such as nonlinearity,[9,10] dynamics,[11−13] time-varying system,[14] missing data,[15−17] multiple data blocks,[18−20] and multiple operating modes or multiple grades of products.[21,22] These works made MSPC methods feasible to be applied on real industrial processes.[23−25]

By performing PCA, a two-way data matrix $X(n \times m)$ is decomposed as

$$X = \sum_{j=1}^{m} \boldsymbol{t}_j \boldsymbol{p}_j^{\mathrm{T}} = \boldsymbol{t}_1 \boldsymbol{p}_1^{\mathrm{T}} + \boldsymbol{t}_2 \boldsymbol{p}_2^{\mathrm{T}} + \cdots + \boldsymbol{t}_m \boldsymbol{p}_m^{\mathrm{T}} \qquad (14.1)$$

where n is usually the number of samples; m is the number of variables; $\boldsymbol{t}_j(n \times 1)$ are latent vectors, which are also named as principal component (PC) vectors or score vectors; $\boldsymbol{p}_j(m \times 1)$ are loading vectors that transform original process variables into PCs. In the decomposition, scores are made to be orthogonal to each other, which means $\boldsymbol{t}_i^{\mathrm{T}} \boldsymbol{t}_j = 0$ when $i \neq j$. At the same time, loadings are orthonormal, which means $\boldsymbol{p}_i^{\mathrm{T}} \boldsymbol{p}_j = 0$ when $i \neq j$ and $\boldsymbol{p}_i^{\mathrm{T}} \boldsymbol{p}_j = 1$ when $i = j$.

The first PC $\boldsymbol{t}_1$ is defined to have the maximum variance subject to $\|\boldsymbol{p}_1\| = 1$; the second PC $\boldsymbol{t}_2$ has the second maximum variance subject to $\|\boldsymbol{p}_2\| = 1$, and is linear independent of $\boldsymbol{t}_1$; and other PCs are defined in a similar way. Algebraically, $\|\boldsymbol{t}_j\|$ is equal to the jth largest eigenvalue λ_j of the covariance matrix $\Sigma = \frac{X^{\mathrm{T}}X}{n-1}$, and $\boldsymbol{p}_j$ is the corresponding eigenvector. Therefore, it is easy to understand that the first several PCs contain most variance information of X, and the last several PCs may only contain measurement noises. Consequently, as shown in Equation 14.2, most systematic variation information can be extracted by retaining a first few PCs, and the dimensions of variables are largely reduced. Process analysis, monitoring, and fault

diagnosis are significantly simplified by working in the low-dimensional subspaces.

$$X = \sum_{j=1}^{A} \boldsymbol{t}_j \boldsymbol{p}_j^{\mathrm{T}} + \sum_{j=A+1}^{m} \boldsymbol{t}_j \boldsymbol{p}_j^{\mathrm{T}} = T P^{\mathrm{T}} + E = \hat{X} + E \tag{14.2}$$

where A is the number of retained PCs, $T(n \times A)$ is the retaining score matrix that contains systematic variation information, $P(m \times A)$ is the loading matrix that extracts variable correlation information, $\hat{X}(n \times m)$ is the reconstruction of X, and $E(m \times n)$ is the residual matrix.

Different types of algorithms such as singular value decomposition (SVD), nonlinear iterative partial least squares (NIPALS), and so on have been developed for the calculation of loading matrix.[7,8] To determine the proper retained number of PCs, the use of cross-validation and many other procedures was introduced by researchers.[7,8,26]

14.2.2 PCA-Based Process Monitoring and Fault Diagnosis

Through PCA modeling, the original data space is decomposed into two orthogonal subspaces: score space and residual space. After a new sample $\boldsymbol{x}^{\mathrm{T}} = [x_1, \ldots, x_m]$ is measured, the measurements can be projected to the score space using the achieved PCA model. The scores $\boldsymbol{t}$ and the residuals $\boldsymbol{e}$ can be calculated using the equations given below:

$$\boldsymbol{t}^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}} P \tag{14.3}$$

$$\hat{\boldsymbol{x}}^{\mathrm{T}} = \boldsymbol{t}^{\mathrm{T}} P^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}} P P^{\mathrm{T}} \tag{14.4}$$

$$\boldsymbol{e}^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}} - \hat{\boldsymbol{x}}^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}} \left(I - P P^{\mathrm{T}}\right) \tag{14.5}$$

For process monitoring, two statistics are calculated.[7] In score space, Hotelling T^2 statistic summarizes systematic variation information:

$$T^2 = \boldsymbol{t}^{\mathrm{T}} S^{-1} \boldsymbol{t} = \sum_{i=1}^{A} \frac{t_i^2}{\lambda_i} \tag{14.6}$$

where $\boldsymbol{t}^{\mathrm{T}} = [t_1, \ldots, t_A]$, $S = \mathrm{diag}(\lambda_1, \ldots, \lambda_A)$, t_i is the ith score variable, and λ_i is the ith largest eigenvalue of the covariance matrix $\sum$, indicating the estimated variance of t_i. Assuming that each score obeys normal distribution, the control limits of T^2 in normal operation can be calculated using F-distribution[7]:

$$T_\alpha^2 = \frac{A(n-1)}{n-A} F_{A,n-A,\alpha} \tag{14.7}$$

where α is the significant level, and $F_{A,n-A,\alpha}$ is the critical value of F-distribution with significant level of α and degrees of freedom of A and $n–A$, whose value can be found in statistical table.

The squared prediction error (SPE) statistic summarizes residual information as

$$\mathrm{SPE} = \boldsymbol{e}^{\mathrm{T}} \boldsymbol{e} = \sum_{j=1}^{m} \left(x_j - \hat{x}_j\right)^2 \tag{14.8}$$

where x_j is the measurement value of the jth process variable, and $\hat{x}_j$ is the corresponding reconstructed value based on PCA. With the assumption that residuals are normal distributed when there is no fault, the control limits of SPE are derived as[7,27]

$$\mathrm{SPE}_\alpha = \theta_1 \left(\frac{C_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right)^{\frac{1}{h_0}} \tag{14.9}$$

where $\theta_i = \sum_{j=A+1}^{m} \lambda_j^i (i = 1, 2, 3)$, $h_0 = 1 - (2\theta_1\theta_3 / 3\theta_2^2)$, C_α is the critical value of normal distribution under significant level of α.

Nomikos and MacGregor[28] provided another way for the calculation of SPE control limits, which is based on the weighted χ^2 distribution $g\chi_h^2$. The basic idea is that the mean $\mu = gh$ and the variance $\sigma^2 = 2g^2h$ of the $g\chi_h^2$ distribution are equal to the sample mean b and the variance v at each time interval. Therefore, g and h can be estimated as $\hat{g} = v/2b$ and $\hat{h} = 2b^2/v$. Consequently, the control limits of SPE at significance level α can be estimated as

$$\mathrm{SPE}_\alpha = \left(\frac{v}{2b}\right) \chi^2_{2b^2/v,\alpha} \tag{14.10}$$

T^2 and SPE are usually used to detect two different types of faults. Hotelling T^2 statistic summarizes variation information contained in scores and measures the distance between each operation point and historical normal operation region, while SPE is a measurement of model fitness. In on-line and off-line process monitoring, the values of these two statistics are plotted on two control charts and compared with corresponding control limits to check whether the process is in control or not. If the process is detected to be not in control, a fault is captured.

When a fault is detected using T^2 or SPE control chart, the cause of such a fault is desired to know. To do fault diagnosis, contribution plots[29] are the most widely applied tools for showing the contribution of each process variable to such a fault with bar plots. Usually, high contribution values indicate the potential causes of faults.

The contribution of the ith PC t_i to T^2 presented in percentage can be derived as

$$C_{t_i} = \frac{t_i^2}{T^2}, \qquad i = 1, \ldots, A \tag{14.11}$$

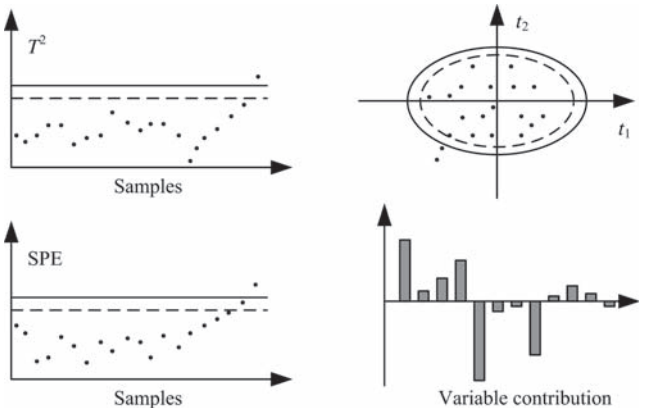

FIGURE 14.2 Typical charts used in MSPC.

And the contribution of variable x_j to t_i is

$$C_{t_i,x_j} = \frac{x_j p_{j,i}}{t_i}, \qquad i = 1, \ldots, A; j = 1, \ldots, m \tag{14.12}$$

where $p_{j,i}$ is the loading for the jth variable on the ith PC.

For SPE, the contribution of variable x_j is

$$C_{\mathrm{SPE},x_j} = \mathrm{sign}(x_j - \hat{x}_j)\frac{(x_j - \hat{x}_j)^2}{\mathrm{SPE}} \tag{14.13}$$

where $\mathrm{sign}(x_j - \hat{x}_j)$ is used to indicate the sign of residuals.

Figure 14.2 shows some typical types of monitoring and diagnosis charts commonly utilized in MSPC.

14.2.3 Normalization

Normalization is a necessary step before performing PCA on a data set.[7] Proper normalization can emphasize correlations among variables, while reducing nonlinearity and eliminating the effects of variable units and measuring ranges. The most common way of normalization includes removing means and equalizing variances, which is sometimes called *autoscaling*. Usually, the first step in autoscaling is named as *centering* while the second is called *scaling*.

In centering, the average value of each variable is subtracted from the original data. For a data matrix $X(n \times m)$, the formula of centering is

$$\tilde{x}_{i,j} = x_{i,j} - \bar{x}_j, \qquad i = 1, \ldots, n; \; j = 1, \ldots, m \tag{14.14}$$

where $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{i,j}$, n is the number of samples, m is the number of variables, i is the sample index, and j is the variable index. As shown in Equation 14.14, centering makes a shift in the data and has no effect on variable correlations or relative positions among samples.

Scaling is often performed after centering. A simple example can illustrate the importance of scaling. Suppose there are two variables in a process, which have same units but different measuring ranges. The measuring range of the first variable is from 0 to 1000 while the other variable changes from 0 to 10. Without scaling, when the measurement values of both variables change with a magnitude of 9, the variations in both variables seem to be same. However, actually, the first variable only varies slightly compared to its measuring range while the second one experiences a rather larger variation. This example shows the effect of measuring range. It is easy to imagine that different variable units can cause similar problems. Therefore, it is necessary to scale the variables to similar ranges before analysis. A common way is transforming all variables' variances to unit. The formula is given as

$$\tilde{x}_{i,j} = \frac{x_{i,j}}{s_j}, \qquad i = 1, \ldots, n; \; j = 1, \ldots, m \tag{14.15}$$

where $s_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{i,j} - \bar{x}_j)^2}$

The uniform equation of autoscaling normalization including both centering and scaling is

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \bar{x}_j}{s_j}, \qquad i = 1, \ldots, n; \; j = 1, \ldots, m \tag{14.16}$$

Besides autoscaling, other normalization methods may also be performed when there is some prior process knowledge. For instance, the important process variables may be scaled to have larger variances than the less important variables.

14.3 MSPC FOR BATCH PROCESSES

For better detection and identification of the abnormal behaviors in batch processes, MSPC methods for batch process monitoring and fault diagnosis have become hot research topics in recent years. Unlike the data of continuous processes, the data collected from a batch process are usually represented by a three-dimensional data matrix $\underline{X}(I \times J \times K)$, where I is the number of total batches, J is the number of process variables, and K is the number of total sampling time intervals in a batch. As PCA can only deal with two-dimensional data matrices, $\underline{X}$ should be unfolded to a two-dimensional data matrix or split into several two-dimensional data matrices before modeling.

To arrange batch data into $\underline{X}$, each run of a batch process should be of equal operation durations. However, in real industries, the batch operation durations of many processes are not fixed in length because of disturbances and changes in operating conditions or control objectives.[30]

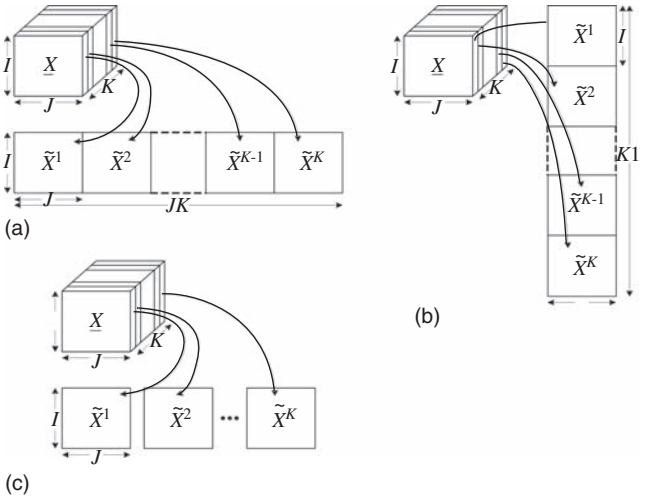

FIGURE 14.3 Unfolding and splitting of three-way batch process data.

In such situations, trajectory alignment, which is also called *synchronization*, is often applied.

In general, for many batch processes, multiple operation phases exist.[31] Evidently, the nature of process of a multiphase batch process is usually different from one phase to another. Meanwhile, multiphase is also an important cause of batch process nonlinearities. Such batch process characteristics should be taken into consideration for more accurate modeling and more efficient monitoring.

The most common way to transform a three-dimensional batch data matrix to a two-dimensional data matrix is batch-wise unfolding. As show in Figure 14.3a, batch-wise unfolding retains the dimension of batches, and merges variable and time dimensions, where $\tilde{X}^k$ is the time-slice data matrix of the kth sampling time. Each row of the unfolded matrix $X(I \times KJ)$ contains all data within a batch. After unfolding, normalization can be performed on this two-way matrix. The mean trajectories of all variables are removed from the data. By doing so, the variations between batches are highlighted. Normalization based on batch-wise unfolding can be called *batch-wise normalization*. Unlike batch-wise unfolding, variable-wise unfolding retains the dimension of variables, and merges the other two dimensions. Each sampling point of each batch is considered as an object, as shown in Figure 14.3b. The process data in the unfolded matrix $X(KI \times J)$ can then be normalized to zero mean and unit variance. After normalization, the grand mean of each variable is eliminated and the information of variable trajectories is left in the data matrix. Such normalization can be called *variable-wise normalization*. Splitting is an alternative choice for batch data pretreatment before PCA modeling.[32] In Figure 14.3c, each two-dimensional matrix represents the data in a time slice from all batches.

On the basis of batch-wise normalization of the batch process data, multiway principal component analysis (MPCA), the most well-known MSPC method for batch process monitoring and fault diagnosis, was proposed.[28,33] The basic idea of MPCA is to perform PCA on the unfolded data matrix $X(I \times KJ)$, supposing X has been normalized. As the whole batch data are regarded as one single object in modeling, both variable cross-correlations and autocorrelations within batch duration are extracted by MPCA loading matrix. Therefore, MPCA works well in off-line monitoring of an entire batch. However, for the same reason, a major drawback of MPCA is that the future unavailable measurements need to be estimated in their on-line applications. The accuracy of such estimation affects the monitoring efficiency of MPCA. Meanwhile, MPCA cannot reveal different phase characteristics. A different type of MPCA model is based on variable-wise unfolded and normalized data matrix $X(KI \times J)$.[34] An advantage of this model is that no estimation of future measurements is necessary in on-line applications. However, as variable-wise unfolding does not focus on the systematic variations around normal trajectories, the monitoring efficiency based on this method is limited.

Other typical batch process MSPC methods include adaptive hierarchical principal component analysis (AHPCA),[35] batch dynamic principal component analysis (BDPCA)/batch dynamic partial least squares (BDPLS)[36] and so on.[32,37−40] However, each of these methods has some limitations. The application of AHPCA is limited by its large computation and storage burdens. BDPCA/BDPLS assumes that the process dynamics structures remain the same throughout the whole batch, ignoring the fact of multiphase. These methods are not liable to be directly applied to injection molding processes. To achieve better monitoring and diagnosis results, the process characteristics of injection molding should be taken into consideration.

14.4 MSPC FOR INJECTION MOLDING PROCESS

As mentioned in Chapter 2, an injection molding cycle includes the following important phases: filling, packing/holding, and cooling. So, it is a typical multiphase batch process. In injection molding, process abnormalities occur owing to various process malfunctions, drifting of process conditions, changes in materials, and other unknown disturbances, leading to the off-spec products. In the meantime, process variables are highly correlated and time varying. Proper designed MSPC methods are necessary for injection molding process modeling, monitoring, and fault diagnosis. As the different operation phases can lead to different process behaviors, which should be treated

with great caution, multiphase MSPC methods[31] should be performed.

14.4.1 Phase-Based Sub-PCA

In the MSPC of multiphase batch processes, phase-based sub-PCA is a preferred method owing to its abilities of identifying phases without prior knowledge and handling the issue of uneven operation durations. The basic idea of sub-PCA is to divide a batch process into several phases and model each phase separately.[41]

The basis of sub-PCA model building contains two levels. First, in multiphase batch processes, different phase can have different variable correlations. Therefore, changes in the correlation structure reflect the changes in the nature of the process, and indicate phase changes. Thus, automatic phase recognition can be performed according to the analysis results of variable correlations. Secondly, within each modeling phase, the nature of process correlation remains the same, although the process may be time varying. Therefore, a series of sub-PCA models can be built for different modeling phases.

There are several major steps in sub-PCA modeling procedure, which are batch process data matrix splitting, phase division, and sub-PCA model building.

As mentioned above, different phases can be indicated by changes in the correlation structure along the direction of time in a certain batch. It is noticed that, in PCA, the loading matrix extracts variable correlation information. For a batch process data matrix $\underline{X}(I \times J \times K)$, each vertical slice $\tilde{X}^k(I \times J)$ is a time-slice data matrix. By performing PCA algorithm on these split time-slice matrices, variable correlation information on each time interval is contained in K number of time-slice loading matrices $\tilde{P}^k$. In the same phase, the time-slice loading matrices are same while different phases have different loading matrices, reflecting changes in the underlying process behavior. The formula of time-slice PCA decomposition is as follows:

$$\tilde{X}^k = \tilde{T}^k(\tilde{P}^k)^{\mathrm{T}}, \qquad k = 1, 2, \ldots, K \tag{14.17}$$

where $\tilde{T}^k$ is the time-slice score matrix containing all J number of score vectors, while $\tilde{P}^k$ is the corresponding loading matrix. Here, the time-slice PCA models are used for phase division instead of on-line monitoring, so it is not necessary to divide score space and residual space. Consequently, all scores are retained in the model.

As each column of a PCA loading matrix contains different amount of process variation information, the time-slice loading matrices $\tilde{P}^k$ are transformed into a weighted form with the importance of each column taken into consideration:

$$\tilde{P}^k = \left[\boldsymbol{p}_1^k \cdot g_1^k, \boldsymbol{p}_2^k \cdot g_2^k, \ldots, \boldsymbol{p}_j^k \cdot g_j^k\right] \tag{14.18}$$

where $\boldsymbol{p}_j^k$ is the jth column of $\tilde{P}^k$, $g_j^k = \lambda_j^k / \sum_{i=1}^{J} \lambda_i^k$, and λ_j^k is the eigenvalue of the covariance matrix $\frac{(\tilde{X}^k)^{\mathrm{T}}\tilde{X}^k}{I-1}$.

Then the phase-division result can be achieved by comparing all the time-slice weighted loading matrices. Such comparison is carried on with a variant k-means clustering algorithm.[42] Each time-slice weighted loading matrix is regarded as a pattern to be clustered, and the Euclidean distance between two patterns is used to assess the dissimilarity. An important parameter in clustering is the threshold of the minimal distance between two cluster centers or the maximal distance between two clusters. With such threshold, k-means clustering algorithm automatically determines the optimal number of final clusters by minimizing the local squared error (for patterns within each cluster) and the global squared error (for all the patterns). Larger threshold leads to fewer clusters and less accurate modeling. A step is added to the clustering algorithm to eliminate singular clusters that catch few patterns in the iterative clustering procedure to enhance the robustness and reliability of the partition algorithm. This improved k-means clustering algorithm groups the time slices with similar correlation structures. As these time-slice loading matrices are extracted along the sampling time of a batch process, the clustering result can be directly associated with the operation time, making the partition of the patterns well interpretable. Thus, a phase can be identified as a series of successive samples in the same cluster and the final phase-division results are achieved based on the clustering result in association with operation time information.

After all phases are divided, sub-PCA models for each phase are then built by taking the average of time-slice PCA models in the corresponding phase:

$$P_c^* = \frac{1}{n_c}\sum_{k=1}^{n_c} \tilde{P}_c^k, \qquad c = 1, 2, \ldots, C \tag{14.19}$$

where C is the total number of phases identified, n_c is the number of time slices in phase c, P_c^* is the representative sub-PCA loading matrix for phase c, and $\tilde{P}_c^k$ is the kth time-slice loading matrix in phase c. The singular value diagonal matrix S_c^* of phase c can be defined in a way similar to time-slice loading matrix:

$$S_c^* = \frac{1}{n_c}\sum_{k=1}^{n_c} \tilde{S}_c^k = \mathrm{diag}\left(\lambda_1^{*c}, \lambda_2^{*c}, \ldots, \lambda_J^{*c}\right), \qquad c = 1, 2, \ldots, C \tag{14.20}$$

where $\tilde{S}_c^k = \mathrm{diag}(\lambda_1^k, \lambda_2^k, \ldots, \lambda_J^k)$ is the kth time-slice singular value diagonal matrix in phase c.

For process monitoring, the number of retained PCs A of each phase sub-PCA model should be calculated. The

cumulative explained variance rate is used as a criterion:

$$A = \min_{A} \left(\sum_{i=1}^{A} \frac{\lambda_i^{*c}}{\text{trace}(S_c^*)} \geq 90\% \right) \tag{14.21}$$

Correspondingly, each P_c^* can be divided into two parts, $\bar{P}_c^*$ and $\tilde{P}_c^*$, for score space and residual space, respectively, so does S_c^*. Therefore, for each time-slice data matrix in phase c,

$$\tilde{T}^k = \tilde{X}^k(\bar{P}_c^*) \tag{14.22}$$

$$\hat{\tilde{X}}^k = \tilde{T}^k(\bar{P}_c^*)^{\mathrm{T}} \tag{14.23}$$

$$\tilde{E}^k = \tilde{X}^k - \hat{\tilde{X}}^k = \tilde{X}^k \left(I - \bar{P}_c^*(\bar{P}_c^*)^{\mathrm{T}}\right) \tag{14.24}$$

Then for on-line monitoring, T^2 and SPE statistics and the corresponding control limits can be calculated based on the models, as is done in conventional PCA.

The major steps of sub-PCA modeling are summarized in Figure 14.4.

For on-line monitoring, before calling a phase model, the current phase should be determined first. Since the phases have been associated with particular time span in the phase-division step, the phase which the new measurements belong to could be easily found by checking the current time interval. Then, the sub-PCA model of the corresponding phase is used to monitor the on-line process data. If there is a fault detected by T^2 or SPE, contribution plots are used to find the reason of the fault as previously described.

The procedure of sub-PCA-based on-line batch process monitoring is shown in Figure 14.5.

Compared with MPCA that models an entire batch process with a single model, sub-PCA models better reflect

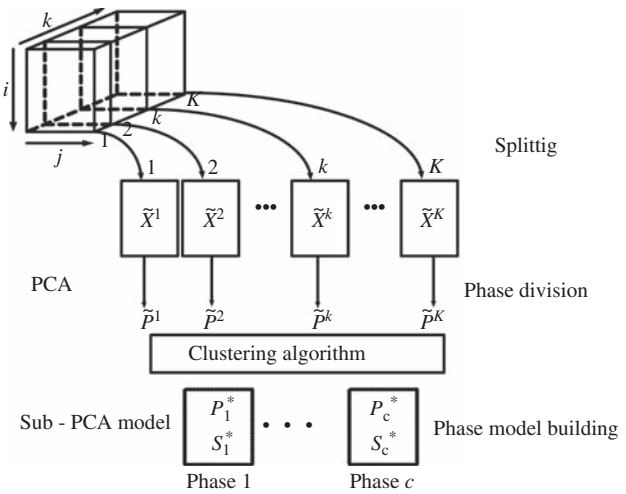

FIGURE 14.4 Illustration of phase division and phase-based sub-PCA modeling.[46]

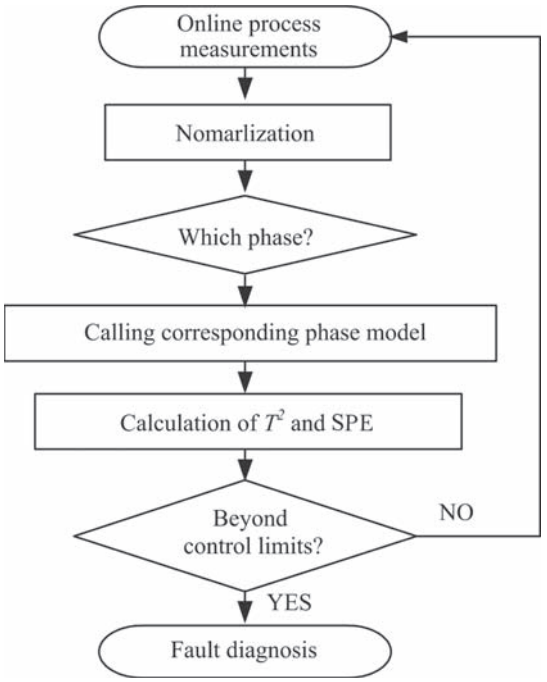

FIGURE 14.5 Procedure of sub-PCA-based on-line batch process monitoring.

the local characteristics of each phase and avoid future measurement estimation in on-line application. The fault detection efficiency can also be improved. A point may be argued is that sub-PCA only models and monitors the cross-correlations among process variables while the process dynamics is ignored. However, the sub-PCA algorithm can be easily modified to take variable autocorrelation (dynamics) information into consideration. This target can be achieved by including the lagging variables into sub-PCA modeling, as dynamic principal component analysis (DPCA)[43] does.

14.4.2 Sub-PCA for Batch Processes with Uneven Operation Durations

Most MSPC methods for batch process monitoring and quality prediction are based on the assumption that the batch durations are same, which means that the data used for modeling should be of equal length and arranged into a three-dimensional matrix $\underline{X}(I \times J \times K)$. However, real industrial processes often have different batch durations from run to run because of disturbances in operating conditions. In multiphase batch processes, not only the whole batch durations but also the durations in each phase can be different from run to run due to disturbances in operating conditions or different process settings. Although different types of batch trajectory synchronization methods have been proposed,[28,30,44,45] shortcomings still exist and

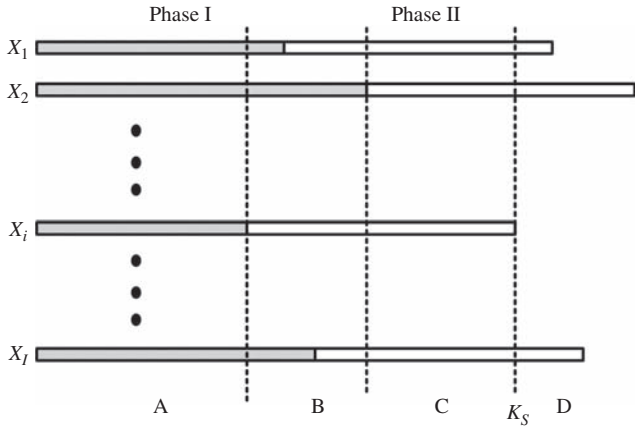

FIGURE 14.6 Illustration of an uneven-length batch process.[46]

these methods do not consider the multiple phase features.[31] Therefore, a method to model and monitor multiphase batch processes without trajectory synchronization is desired. The phase-based sub-PCA method was extended to uneven-duration batch processes to solve the problem.[46]

The complexity of the data structure of uneven-duration multiphase batch processes is illustrated with a two-phase batch process plotted in Figure 14.6. In this figure, X_i are the data from the ith batch, and K_s is the number of sampling intervals in the shortest batch. Spans A and C are the common part of phases I and II. In span B, some batches are in phase I, while others are in phase II. Span D has incomplete data structure of phase II.

To apply phase-based sub-PCA on multiphase uneven-duration batch processes, the first difficulty to overcome is the way of data normalization. Because of the variations in phase durations, batch-wise normalization is not proper. The mean trajectories and standard deviations at each sampling interval are difficult to be calculated, especially in span D that has an irregular tail. Another problem is about the normalization of the time slices in space B. The time-slice covariance matrix in this span is meaningless because different operation cycles may run in different phases. Therefore, subtracting the mean trajectories and dividing by the standard deviations for the data in this span is meaningless either.

Instead, the variable-wise unfolding and normalization method can be applied directly to uneven-length data under such a situation. However, this kind of unfolding debases monitoring efficiency because it focuses on the variations in variable trajectories along the time direction instead of the variations around the normal trajectories.

To solve the problem discussed above, the variable-wise unfolding and normalization is used for the purpose of phase division. A phase-division model is built for each phase based on such normalization. Then, after phase division, process data are normalized again in the batch-wise way to rebuild sub-PCA models for the on-line monitoring. The details are provided in the following paragraphs.

Without losing generality, the shortest batch length K_s in all reference batches is supposed to be known. After variable-wise unfolding and normalization, K_s time-slice PCA models are built based on time-slice data matrices. Then, the k-means clustering algorithm is performed on the weighted loading matrices of these PCA models as what were introduced in the previous section. As the stable clusters could be found only for the common part of each phase, such as spans A and C in Figure 14.6, the first stable cluster indicates the common part of the first phase. Therefore, a phase-division PCA model can be built for phase I based on this cluster. After selecting the number of retained PCs, the SPE values can be computed. As the process correlation structure should not change significantly until the process enters the next phase, the data in phase I should be explained well by this phase-division PCA model, with relatively small SPE values. In contrast, the model will result in large SPE values for data not belonging to phase I. Thus, the phase-division PCA model can be used to find the actual durations of this phase for all reference batches by checking the SPE values and comparing them with a threshold SPE^*. It is suggested to assign SPE^* with a value that is slightly larger than the maximum SPE values of the samples belonging to the common part of phase I. After the lengths of phase I are identified, the first phase data are removed from the reference batch data, and a new reference data set is formed with the remaining data. The same procedure is repeated to determine the durations of phase II. Such phase-division steps are conducted iteratively to get information of all phases.

After the above-mentioned steps, the data in each phase are renormalized in a modified batch-wise way to achieve better monitoring efficiency. For the data in the irregular tail regions of each phase, the average trajectories and standard deviations are calculated from the available data and then used in normalization. Then, the time-slice PCA models can be built based on such normalization. The time-slice loading matrices for the data in the irregular tail region are calculated from the available data of the longer batches. These time-slice loading matrices are summarized into a phase sub-PCA model for on-line monitoring. The contribution of each time-slice PCA to the phase model is determined with a weight w^k:

$$w^k = I^k / I \tag{14.25}$$

where I^k is the number of available batches for calculating the kth time-slice PCA model and I is the total number of reference batches. The final sub-PCA models for

monitoring are defined as a weighted mean of the time-slice loading matrices:

$$P_c^* = \frac{\sum_{k=1}^{L_c} w^k \tilde{P}_c^k}{\sum_{k=1}^{L_c} w^k}, \qquad c = 1, 2, \ldots, C; k = 1, 2, \ldots, L^c \tag{14.26}$$

where C is the total number of phases identified, L^c is the number of time slices in longest phase c in the reference data set, P_c^* is the sub-PCA loading matrix for phase c to be used in process monitoring, and $\tilde{P}_c^k$ is the kth time-slice loading matrix in phase c. After the determination of the number of retained PCs, T^2 and SPE control limits for each sampling interval can be calculated for on-line monitoring, following the same procedure as is done in the conventional sub-PCA method.

Figure 14.7 shows the procedure of phase division and phase-based sub-PCA modeling for batch processes with uneven operation durations.

A problem to be solved in on-line monitoring is how to differentiate process abnormalities from phase changes as both situations can make the statistics beyond the control limits. Suppose the lengths of phase c are varied from $L^c_{\min}$ to $L^c_{\max}$. For the data belonging to $[1, L^c_{\min}]$, the on-line monitoring can be performed following the monitoring procedure based on the conventional sub-PCA. If the data belong to $(L^c_{\min}, L^c_{\max}]$, there are two possibilities if the SPE or T^2 values are outside of the control limits: a fault or process entering the next phase. When such a situation arises, the data are renormalized and monitored with the monitoring model of phase $c+1$. If the statistics exceed the control limits again, a fault that has occurred is detected. Otherwise, it indicates the beginning of a new phase.

The procedure of sub-PCA-based uneven-length batch process on-line monitoring is illustrated in Figure 14.8.

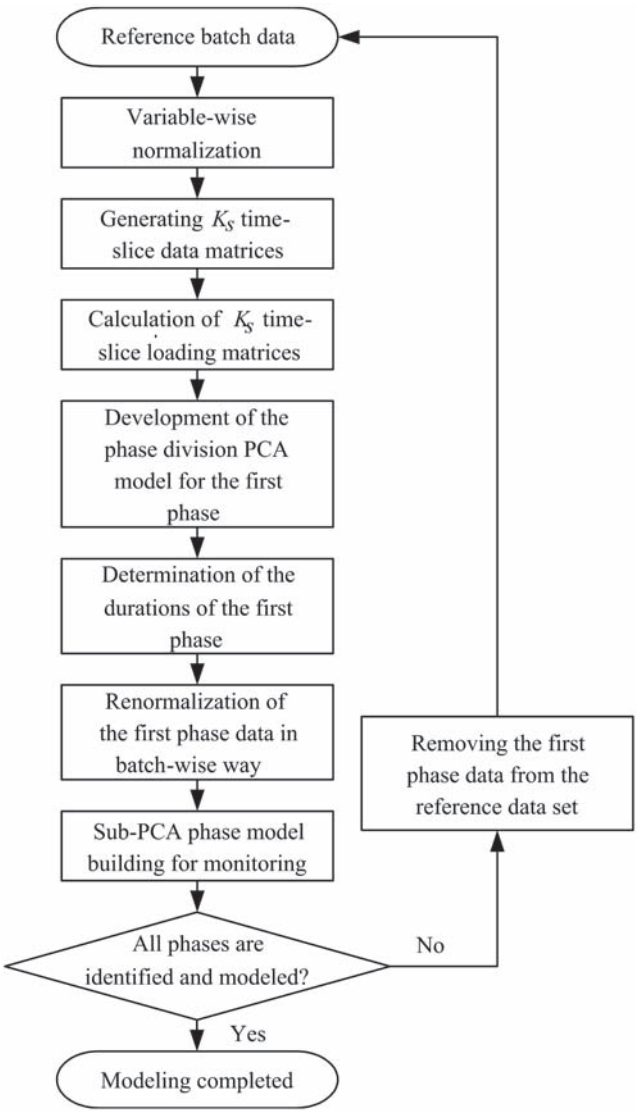

FIGURE 14.7 Procedure of phase division and phase-based sub-PCA modeling for uneven-duration batch process data.

14.4.3 Sub-PCA with Limited Reference Data

Usually, the modeling methods based on MSPC require large amounts of normal historical data that cover the entire normal operating region. However, owing to the rapid changing requirements of the commercial market, the operation conditions change frequently. As a result, to collect a large number of reference data under each condition may be too time consuming. If there is a method to model such processes with limited data, this problem could be solved. With such a motivation, a phase-based batch process monitoring method with limited reference data was proposed.[47,48]

The major difference between this method and the conventional sub-PCA is that a moving window is used to extract the local process characteristics instead of the time-slice data matrix. In each window, data are arranged as a two-way matrix, whose rows contain the values of process variables measured at different sampling intervals along the time direction in the window.

In this method, phase division and modeling starts with the data from an arbitrary batch in normal operation. The data of this batch is stored as $X(K \times J)$, where K is the number of sampling intervals in a batch and J is the number of process variables. The moving window strategy is used along the time direction in this cycle. Therefore, the data in each window form a two-way matrix $\tilde{X}^k(d \times J)$, where k is the index of window and d is the length of a window. The moving step can be simply set to 1. Thus, there are $(K - d + 1)$ number of windows formed totally, as shown in Figure 14.9.

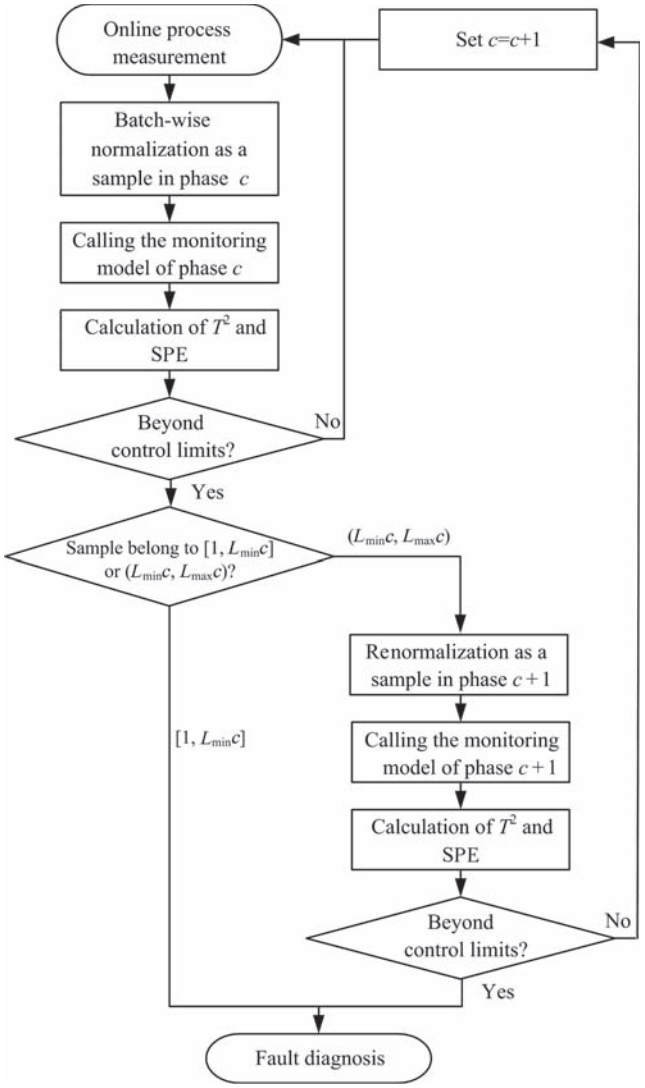

FIGURE 14.8 Procedure of sub-PCA-based on-line monitoring for uneven-duration batch processes.

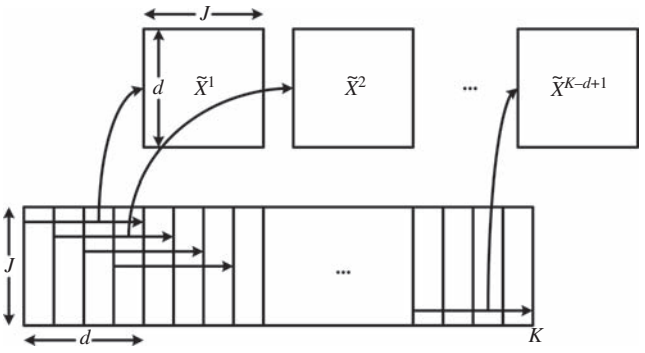

FIGURE 14.9 Moving data window for one batch data.

Then, the data in each window are normalized as follows:

$$\begin{cases} \dfrac{x_j - \bar{x}_j}{s_j} & \text{if } s_j \geq \varepsilon \\ (x_j - \bar{x}_j)s_j & \text{if } s_j < \varepsilon \end{cases} \tag{14.27}$$

where j is a variable index, $\bar{x}_j$ is the mean value of x_j in the window, and ε is a small value smaller than 1. In each moving window, the centering can eliminate a large part of nonlinearity caused by process dynamics as the process variable trajectories do not vary greatly. The purpose of doing scaling is to reduce the influences of the variables of little variation in the window. Therefore, the effects of random noise variations will not be amplified.

A PCA model is then built for the data in each window and the local correlation information is extracted based on the loading matrix. Then, the phase division can be conducted with clustering of these window-based PCA loading matrices similar to the procedure in the conventional sub-PCA method.

In each window, a PCA model is derived after data normalization with which the local variable correlation structure information is extracted. Consequently, switch points between phases can be identified according to the changes in process correlation structures. The k-means clustering algorithm is adopted to divide the $(K - d + 1)$ number of loading matrices into groups. Then, the phase division results are achieved based on the clustering results associated with batch operation time.

After phase division, sub-PCA models can be built to monitor each phase by taking the averages of the window PCA models in the corresponding phase. The number of PCs is determined using the cumulative explained variance rates. Then, the control limits of T^2 and SPE can be calculated accordingly for on-line monitoring.

As mentioned above, the control limits are initially calculated based on the data of one reference batch. When more normal operation data are available, control limit update can be performed in order to focus more on batch-to-batch variation. In detail, the data in each window is gradually filled with the new data:

$$\tilde{X}_{i+1}^{k} = \begin{pmatrix} \tilde{X}_{i}^{k} \\ x_{i+1}^{k} \end{pmatrix} \tag{14.28}$$

where k is the index of window, i is the index of the available batch number, $\boldsymbol{x}_{i+1}^{k}$ is the new batch data in the kth window, $\tilde{X}_{i}^{k}$ is the data matrix in the kth window after the data i batches are available. Based on the updated process data matrices, the control limits of the T^2 and SPE statistics can be recalculated. Therefore, the model precision is improved further, and better monitoring efficiency is hopefully achieved.

The on-line monitoring procedure of this method is similar to the one based on the conventional sub-PCA.

This method was revised later by using several reference batches instead of one batch in initial batch process modeling in order to cover more process variance information in both time and batch directions. Both the phase PCA models and the control limits are updated when more normal data are available.

Suppose data from I batches are used in phase division and initial modeling. For each operation cycle, the batch data are stored as $X(K \times J)$. The moving window covers I batches and d sampling intervals from each cycle. Therefore, the data in each window form a two-way matrix $\tilde{X}^k(Id \times J)$, and there are totally $(K - d + 1)$ number of windows. Such moving window is named as *generalized moving window*.

In each generalized window, the process data are normalized to have means of zero and unit variances. The following steps of window PCA calculation, phase division based on clustering, phase PCA modeling, and initial control limits computation follow the same procedure as mentioned earlier.

During on-line monitoring, the phase models are updated to contain information of new normal batches when data from several successful new batches are available. As each phase has different process behaviors and correlation structures, different phase models are updated with different updating ratios. The concept of process data dissimilarity[49] is used and a forgetting factor is calculated for each phase model updating. A classification method based on the Karhunen–Loeve (KL) expansion[50] is adopted to evaluate the difference between the old reference batches distribution and the current new modeling batches distribution. The calculated dissimilarity index D is then used as a forgetting factor to indicate how much new information should be included into and how much old information should be excluded from the phase model $\bar{P}_c^*$. When new batches are similar to old batches, D is close to zero and the updated phase model is similar to the old one. On the contrary, if D is close to one, the similarity between new normal batches and old batches is small and the model should be updated with a high updating ratio:

$$\bar{P}_{c,\text{cur}}^* = (1 - D_{c,\text{cur}}) \times \bar{P}_{c,\text{pre}}^* + D_{c,\text{cur}} \times \bar{P}_{c,\text{upd}}^* \quad (14.29)$$

where $\bar{P}_{c,\text{cur}}^*$ is the current updated monitoring model for the cth phase, $\bar{P}_{c,\text{pre}}^*$ is the previous model for this phase, $\bar{P}_{c,\text{upd}}^*$ is the updated model representing the new process distribution, and $D_{c,\text{cur}}$ is the dissimilarity index for phase c. When the number of retained PCs in $\bar{P}_{c,\text{upd}}^*$ is different from that in $\bar{P}_{c,\text{pre}}^*$, the bigger one is chosen as the number of retained PCs in $\bar{P}_{c,\text{cur}}^*$.

The update frequency is also important for modeling and monitoring. It can be measured with an index h which is the number of new modeling batches. A large h results in slow model updating and reflects the process changes from batch to batch slowly. On the other hand, a small h may be too sensitive. It is suggested that the value of h should not be smaller than the number of initial reference batches.

Besides the model parameter, the control limits also need to be updated. In each window, the data used for control limits calculation are augmented with the data from new normal batches.

$$\tilde{X}_{\text{cur}}^k = \begin{pmatrix} \tilde{X}_{\text{pre}}^k \\ \tilde{X}_{\text{upd}}^k \end{pmatrix} \quad (14.30)$$

where $\tilde{X}_{\text{cur}}^k$ is the current augmented data in the kth window, $\tilde{X}_{\text{pre}}^k$ is the old data matrix in this window, and $\tilde{X}_{\text{upd}}^k$ is the new available normal batch data that can be used in updating. The updated control limits can be calculated based on the updated data sets and phase models. Data renormalization may be necessary before the control limits are updated when the data mean and standard deviation drift with the operation evolution.

14.4.4 Applications

14.4.4.1 Application of Conventional Sub-PCA The sub-PCA method has been used in the analysis and monitoring of the injection molding process.[51] Data from past successful operations under the following operating conditions were collected as a reference set: injection velocity was 24 mm/s; mold temperature equaled to 25 °C; seven-band barrel temperatures were set to be 200, 200, 200, 200, 180, 160, 120 °C; packing-holding time was fixed to be 3 s with total cycle time around 20 s. Totally, 60 normal batch runs were conducted under such operating conditions. Totally, 16 process variables as listed in Table 14.1 were measured.

TABLE 14.1 The 16 Process Variables Measured in the Experiment

Variable's Description	Unit
Nozzle pressure	bar
Stroke	mm
Injection velocity	mm/s
Injection pressure	bar
Plastication pressure	bar
Injection cylinder pressure	bar
Cavity pressure	bar
Screw rotation speed	rpm
SV1 opening	%
SV2 opening	%
Cavity temperature	°C
Nozzle temperature	°C
Barrel temperature 1	°C
Barrel temperature 2	°C
Barrel temperature 3	°C
Barrel temperature 4	°C

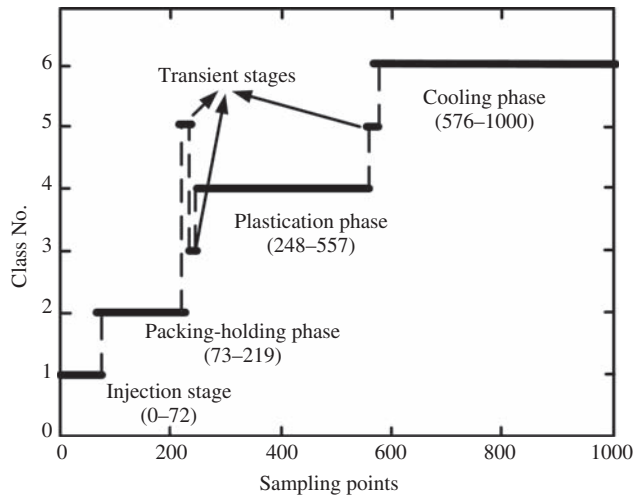

FIGURE 14.10 Phase division of the injection modeling process by sub-PCA modeling.

For a multiphase batch process, different process variables dominate different phases. Therefore, it is desirable to scale process variables within batch run to retain the inherent weights in different phases. Normalization was conducted by mapping the original measurements into [0,1].

The phase-division result based on weighted time-slice loading matrices, after proper data synchronization and normalization, is shown in Figure 14.11. The k-means clustering algorithm divided the process into four major phases. The first two phases corresponded to injection (filling) and packing-holding. The operating phase of cooling was further divided into two modeling phases, where the third phase corresponded to plastication actions and the fourth was cooling phase after plastication was finished. Such phase division results were confirmed by process knowledge and was able to promote the process understanding.

After phase division, sub-PCA models were derived for each phase. The loading plots in Figure 14.11, which plot the second loading vectors against the first ones, show the cross-correlations among process variables in each phase. In each subplot, variables in the same cluster have high correlations among each other. On the contrary, variables in different clusters have weak correlations.[52] In Figure 14.11, process variables form different clusters in different phases, indicating that the variable correlation structure changes from phase to phase. In each subplot, the variables located in the circles are close to the origin of coordinates, indicating their variances are insignificant in the corresponding phase. On the other hand, the variables in the diamonds are important and contribute more to the phase sub-PCA model. All barrel temperature variables (No. 13, 14, 15, 16) are clustered together in the rectangles, which form a cluster independent of other variables, because they were close-loop controlled. The different

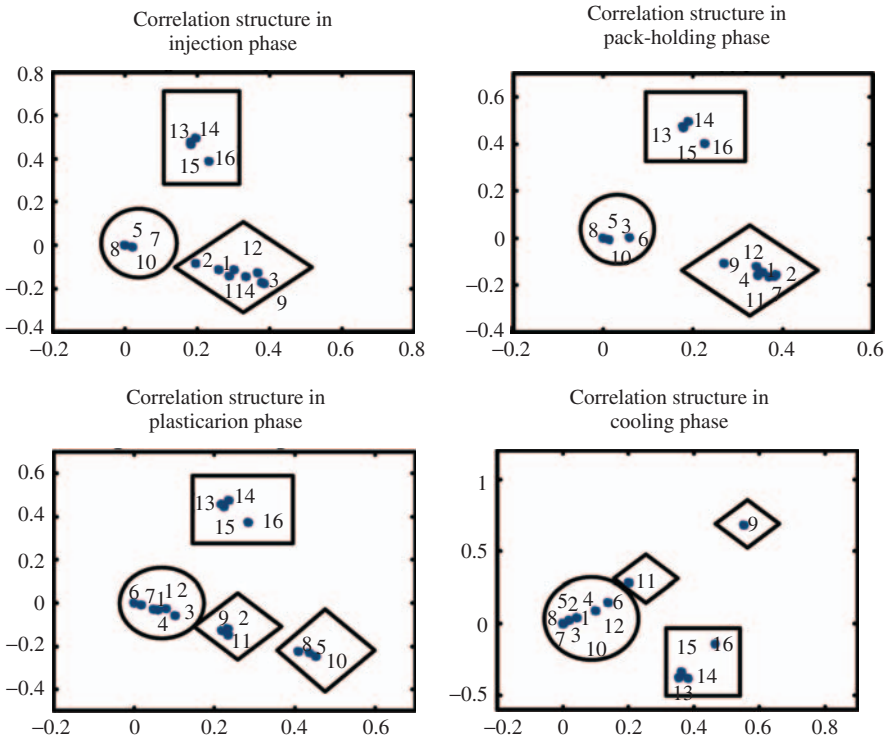

FIGURE 14.11 Correlation structures in the loading plots of sub-PCA models.

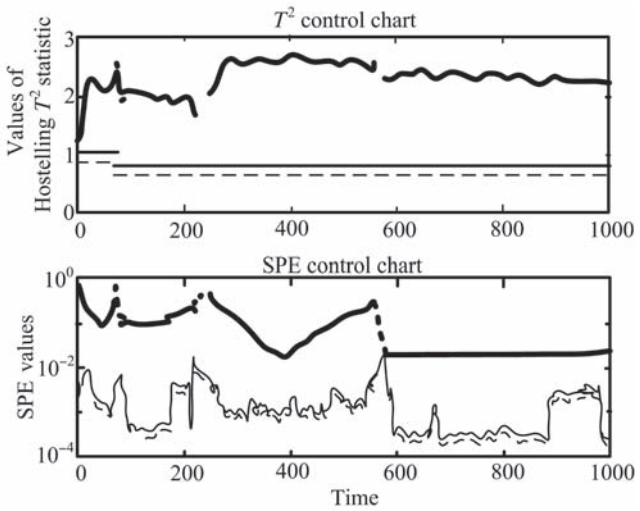

FIGURE 14.12 Monitoring results of material disturbance based on sub-PCA (solid curve: 99% control limits; dash curve: 95% control limits).

phase characteristics indicate that a phase-based process analysis is necessary.

To check the fault detection ability of the sub-PCA method, three typical types of faults were generated.

Fault 1 is due to material disturbance because of adding a small amount of polypropylene (PP) into the original feedstock high density polyethylene (HDPE). The T^2 and SPE control charts were used to detect the fault, as shown in Figure 14.12. The fault was identified soon after the start of the filling phase. As discovered in contribution plots in Figure 14.13, contamination of the feedstock caused a lower cavity temperature (No. 11) in all phases because the PP cools and solidifies faster than HDPE. At the same time, the PP has higher viscosity than HDPE, generating more shear heat and making the nozzle melt temperature (No. 12) higher. In packing-holding phase, the cavity pressure (No. 7) had lower values than that in other phases, which is caused by the faster solidification of PP than HDPE. Such unique phase characteristics can only be revealed by the phase-based method.

The second fault is due to barrel temperature sensor failure. When the fault occurred, the reading of this temperature dropped significantly. This fault was detected efficiently with the control charts shown in Figure 14.14. The drop of the temperature reading resulted in a full heating of this barrel zone. The generated excessive heats were conducted to other zones, making the heating of these zones fully shutdown. Such an analysis was confirmed by the contribution plots shown in Figure 14.14, which shows that the temperature of the failed zone dropped and the temperatures of the neighboring zones increased. Four subplots show similar results because the barrel temperatures form an independent and stable cluster throughout the whole batch.

Fault 3 was caused by a check-ring failure, a common fault in injection molding process. The check-ring valve is a device that allows the polymer melt to flow from

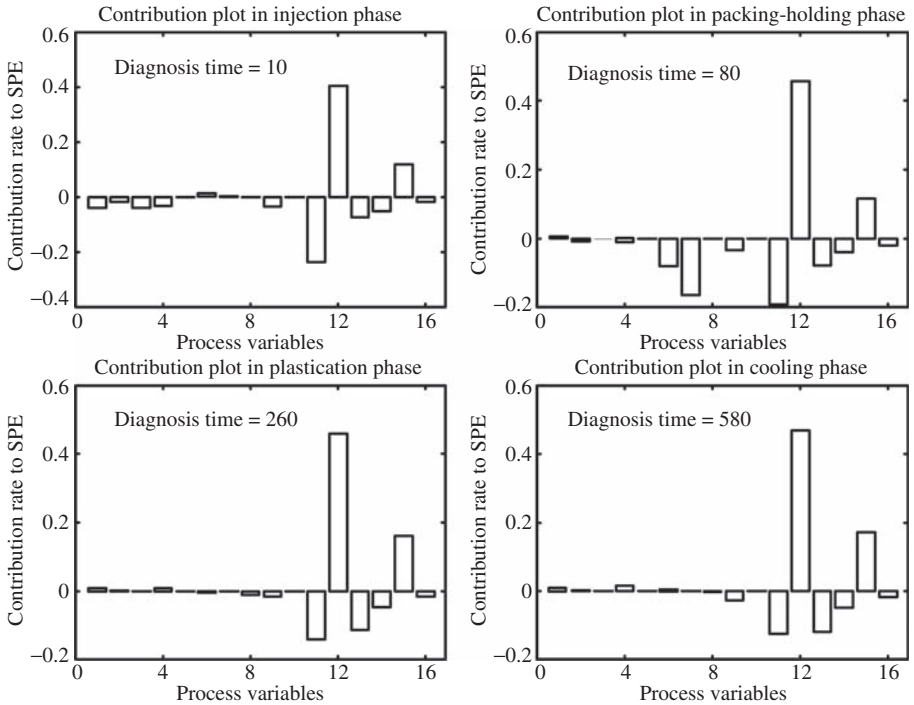

FIGURE 14.13 Contribution plots for a fault due to material disturbance based on sub-PCA.

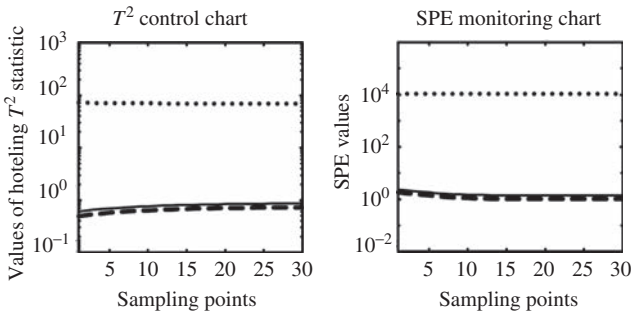

FIGURE 14.14 Monitoring results of a sensor fault based on sub-PCA (solid curve: 99% control limits; dash curve: 95% control limits).

the screw channel to the nozzle during plastication phase and closes to prevent polymer backflow from the nozzle to screw channel during injection and packing-holding phases. The check-ring fault was soon detected by the SPE control chart, and the process variable correlations showed different features in different phases as indicated in Figure 14.16. Owing to the backflow caused by check-ring failure, the amount of material injected into the cavity became smaller in the filling phase. The nozzle pressure (No. 1), injection pressure (No. 4), and cavity pressure (No. 7) all became lower. Therefore, in the next phase (packing-holding), more material was packed into the cavity. Accordingly, the values of stroke (No. 2), screw speed (No. 3), and pressures (No. 4, 5) were all higher than usual. As a longer stroke had moved in the previous two phases, another longer stroke (No. 2) had to be recovered in the plastication phase. All the above-mentioned process characteristics are revealed by the contribution plots shown in Figure 14.16. Such phase-related process knowledge cannot be learned by the methods that take the whole batch as a single object. The different fault patterns shown in different phase contribution plots give a much clearer indication to diagnose the cause of the fault.

14.4.4.2 Application of Sub-PCA on Uneven-Duration Data In an injection molding process, the filling phase finishes when the mold is completely filled, which means that the duration of filling phase is not fixed but depends on the injection velocity. Lower injection velocity leads to longer filling time and longer total batch duration, and vice versa. Under such situations, the sub-PCA method for uneven-duration batch processes should be applied.

In experiments, HDPE was used as the feedstock. Other operating conditions were set as follows: the injection velocity changed from 22 to 26 mm/s to simulate the varying durations in injection phase; the nozzle packing pressure was set at 200 bar; the mold temperature was set at 25 °C; the seven-band barrel temperatures were set at 200, 200, 200, 200, 180, 160, and 120 °C, respectively;

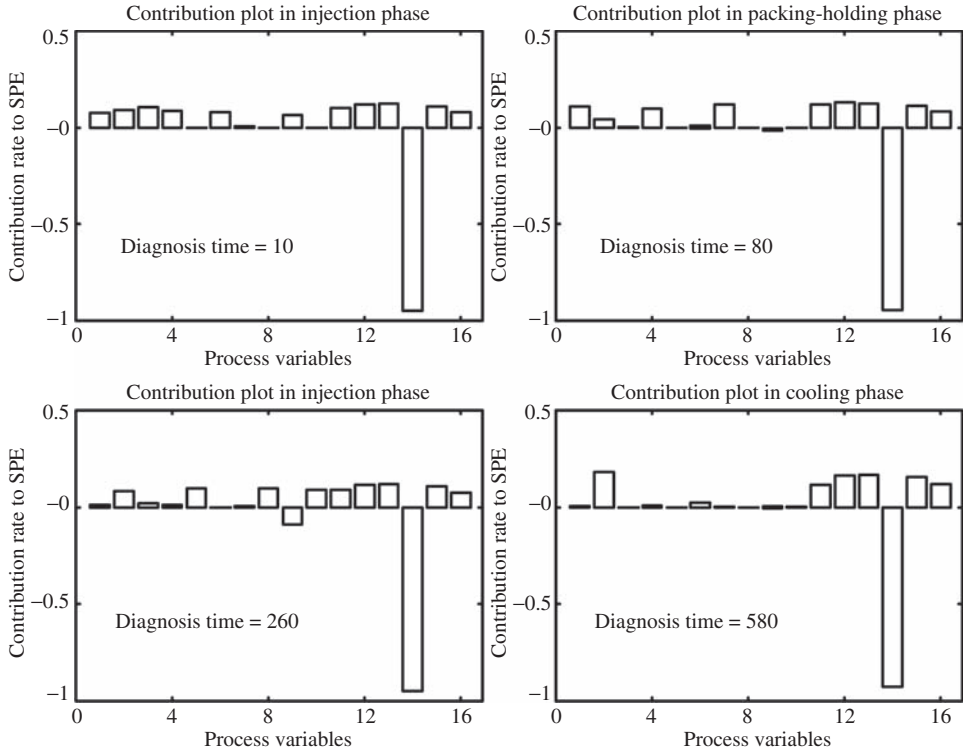

FIGURE 14.14 Contribution plots for the sensor fault based on sub-PCA.

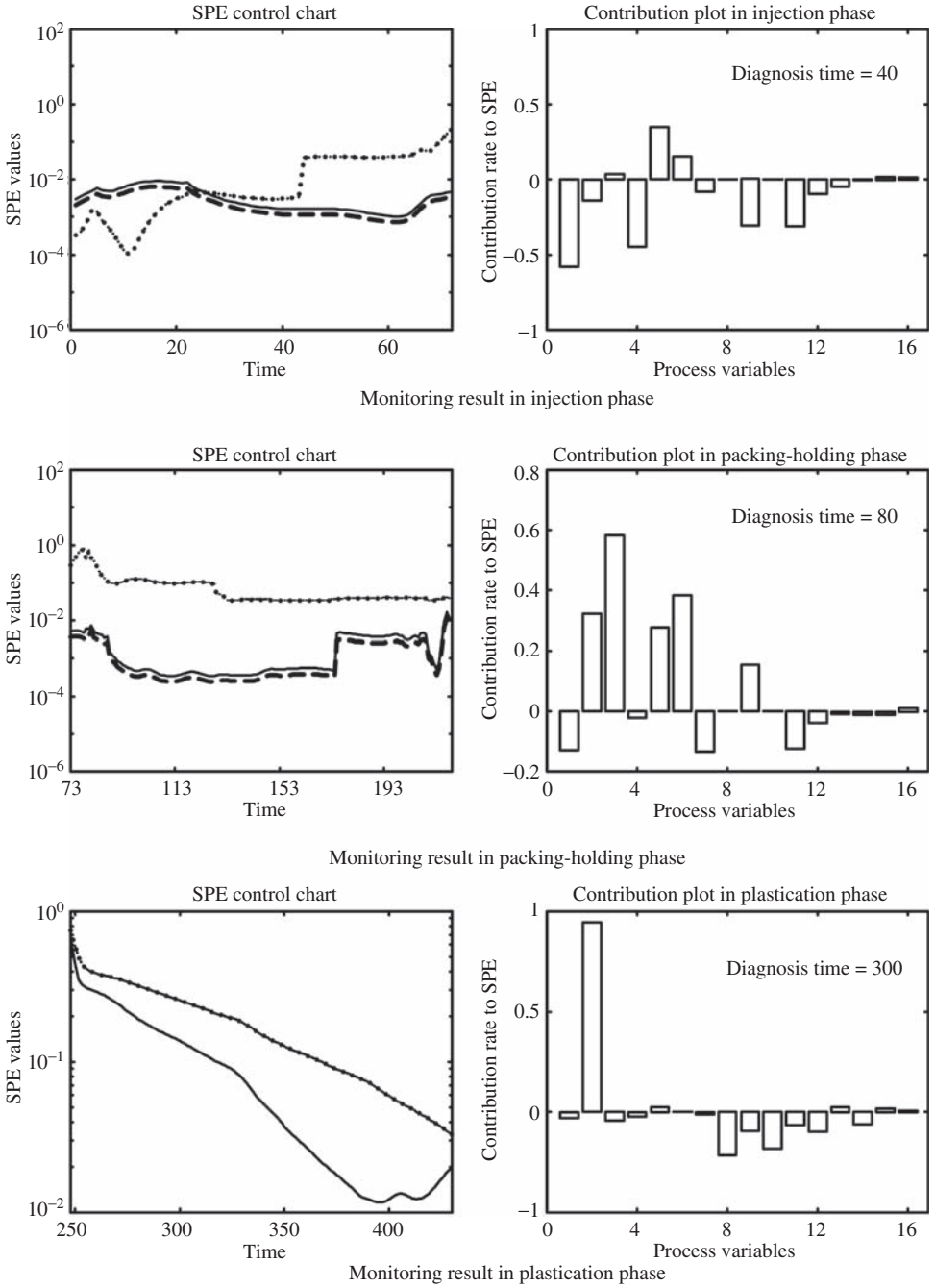

FIGURE 14.16 Monitoring results and contribution plots for a fault due to check-ring problem (solid curve: 99% control limits; dash curve: 95% control limits).

the packing-holding time was fixed at 6 s; the cooling time was set at 15 s; and the sampling interval was 20 ms. Under such settings, the maximum difference in the number of data points in the injection phase was 13, about 15% of the average filling duration, while the other phases were controlled to have exactly the same data length. In total, 35 normal batch runs were conducted, where the shortest and longest batch lengths were 1180 and 1193 samples, respectively. The process variables measured in this application are listed in Table 14.2.[46]

After variable-wise normalization, $K_s (K_s = 1180)$ time-slice PCA loading matrices were calculated and used as the input for the k-means clustering algorithm. The clustering results are shown in Figure 14.17. There are four major stable clusters in the plot, corresponding to filling, packing-holding, plastication, and cooling phases. The odd points

TABLE 14.2 The 11 Process Variables Measured in the Experiment[a]

Variable's Description	Unit
Nozzle temperature	°C
Nozzle pressure	bar
Stroke	mm
Injection velocity	mm/s
Hydraulic pressure	bar
Plastication pressure	bar
Cavity pressure	bar
Screw rotation speed	rpm
SV1 opening	%
SV2 opening	%
Cavity temperature	°C

[a]Ref. 46.

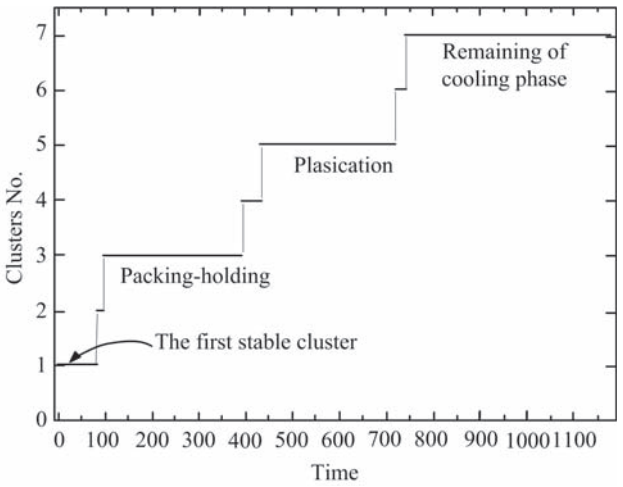

FIGURE 14.17 Clustering results for the determination of the first phase.[46]

between these stable clusters were caused by the uneven durations of the filling phase. The first phase model for phase identification was then built based on the first stable cluster information. The first phase determination results of three normal batches are shown in Figure 14.18. The injection velocities of the test batches were 22, 24, and 26 mm/s, respectively, leading to uneven-length filling phase. The faster injection corresponds to the shorter filling phase, which obeys the process knowledge. In all reference batches, the shortest and longest filling phases had 80 and 93 samples, respectively, indicating the filling phase duration of a normal batch should be within the range of [80, 93].

After removing the filling phase data from the reference set, the remaining data had the same lengths. The above steps were repeated to identify the last three phases. The clustering results are shown in Figure 14.19. The short period between the packing-holding phase and the

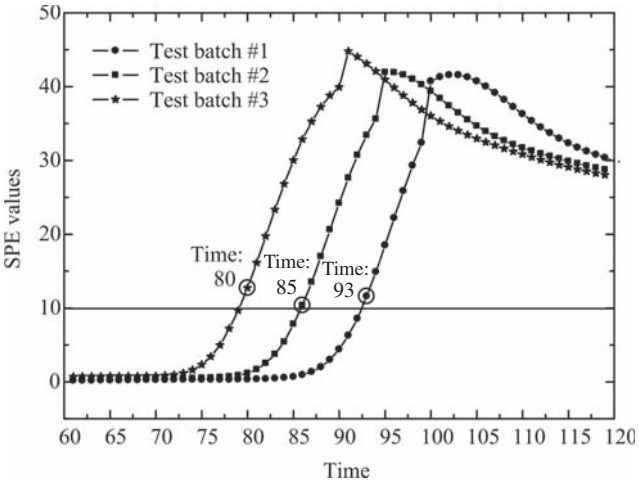

FIGURE 14.18 Determination of the first phase durations for three normal batches.[46]

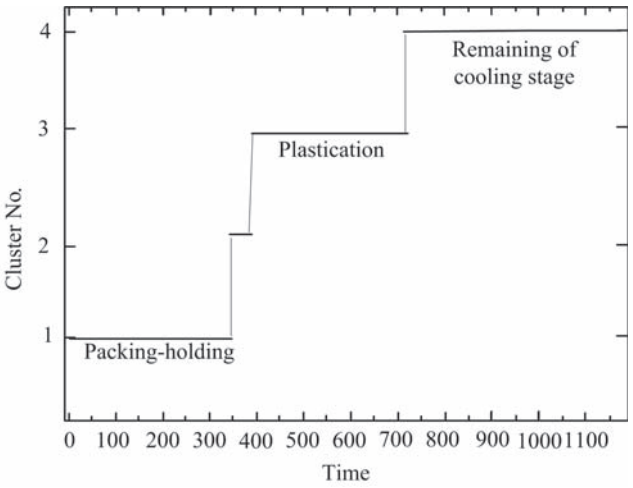

FIGURE 14.19 Clustering results for the determination of other phases.[46]

plastication phase was due to the operation of "suck-back" to retract the screw for a certain distance to prevent melt-drooling.

After phase division, the batch data were renormalized in a batch-wise way and used to build phase-based sub-PCA models for on-line monitoring. As the filling phase duration was known to fall in the range of [80, 93] after phase division and process analysis, the data samples before 80 were normalized and monitored by regarding them as the data in the first phase. The filling phase sub-PCA model was used to compute the T^2 and SPE statistics. If both statistics were within the control limits, the data were in control. Otherwise, abnormalities were detected.

For the data between [80, 93], if their T^2 or SPE statistics were outside the control limits, they were regarded as the

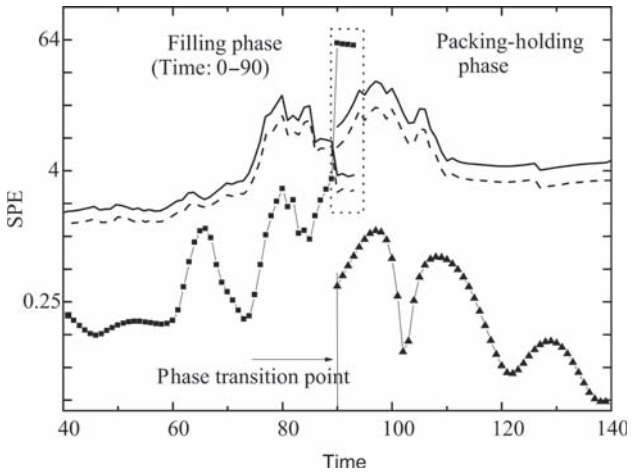

FIGURE 14.20 SPE control chart of a normal batch based on uneven-duration sub-PCA (solid curve: 99% control limits; dash curve: 95% control limits).[46]

beginning of the next phase and renormalized accordingly. The sub-PCA model of the packing-holding phase was then used to monitor such data samples. If the assumption that the process has entered into this phase was true, the T^2 or SPE statistics should both be in control. As shown in Figure 14.20, after the batch operation switched from the filling phase to the packing-holding phase at the 90th sampling interval, the control charts based on the first sub-PCA model gave higher statistic values than control limits. At the same time, the second phase sub-PCA model explained the process well from that perspective.

For the data after sampling interval 93, which was the end of the longest filling phase, the packing-holding phase sub-PCA model was applied to monitor the evolving data of all batches because all normal batches should have finished the filling phase operation. The data in the following phases were monitored in a similar way.

Three kinds of faults, due to check-ring problem, material disturbances, and valve-sticking, were to be detected. The check-ring fault mainly affected the first two phases; the material disturbances changed the variable correlation structures throughout the whole batch; and the valve-sticking happened in the middle of the packing-holding phase. To focus on the uneven-duration problem, only the monitoring and diagnosis results of the first two phases are illustrated in the figures.

The features of first two faults are discussed in the last paragraph. The SPE monitoring charts shown in Figures 14.21a and 14.22a detected the faults clearly and efficiently from the beginning of the filling phase. The contribution plots shown in Figures 14.21b and 14.22b pointed out that the process variables are affected significantly by the abnormalities, which agreed quite well with process knowledge of the fault patterns.

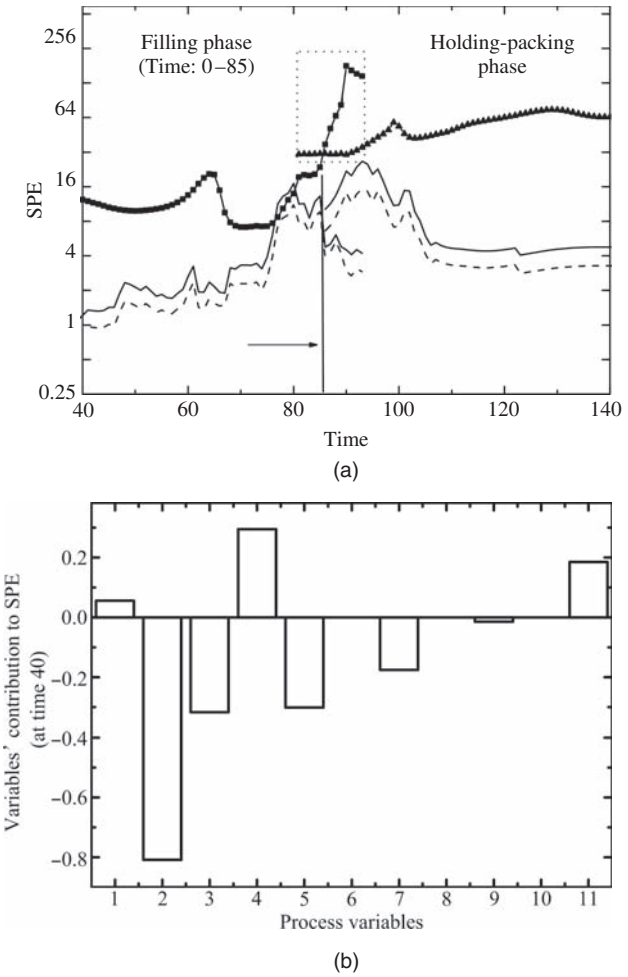

FIGURE 14.21 Monitoring and diagnosis of an abnormal batch with a check-ring fault based on uneven-duration sub-PCA: (a) SPE control chart (solid curve: 99% control limits; dash curve: 95% control limits) and (b) contribution plot.

The SV1 valve-sticking fault caused an uncontrolled nozzle pressure. After the fault occurred in the middle of the packing-holding phase, the nozzle pressure (No. 2) increased. Consequently, the cavity pressure (No. 7) also increased. As the correlations among variables were not changed, the magnitude of SPE statistic did not increase much. However, the value of T^2 statistic became outside the control limits following an increase in the magnitude of the variables, as observed in Figure 14.23a. The contributions of process variables to the first PC are plotted in Figure 14.23b. It is clear that the two pressures mentioned above contributed most to the T^2.

The application results indicate that the uneven-duration sub-PCA method can be applied to uneven-duration injection molding processes successfully. The monitoring efficiency is high. The contribution plots based on this method are quite helpful to dig out knowledge and find the causes of process abnormalities.

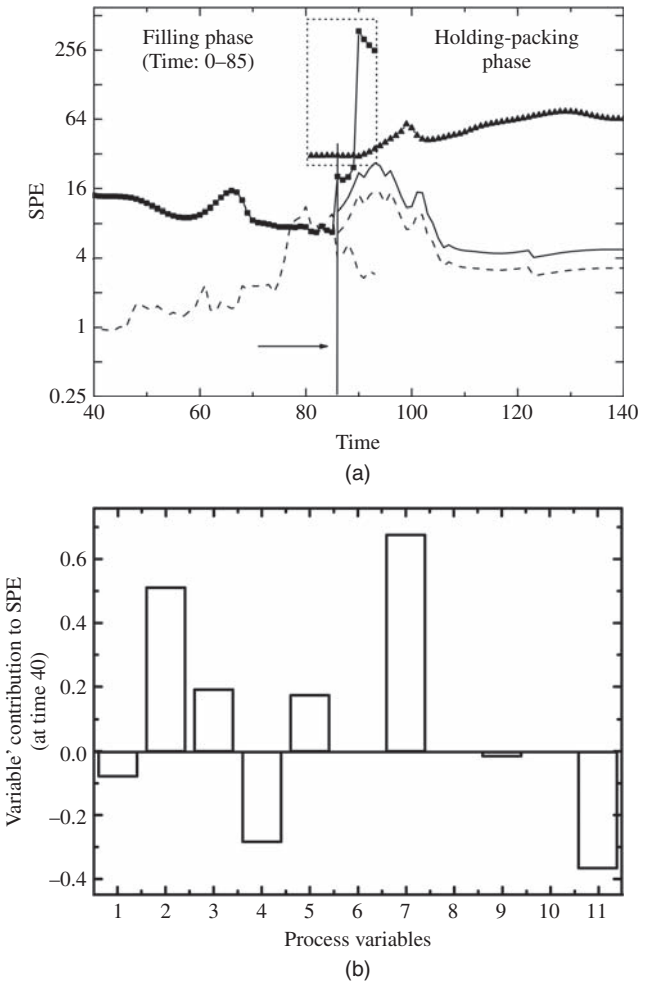

FIGURE 14.22 Monitoring and diagnosis of an abnormal batch with material disturbances based on uneven-duration sub-PCA: (a) SPE control chart (solid curve: 99% control limits; dash curve: 95% control limits) and (b) contribution plot.

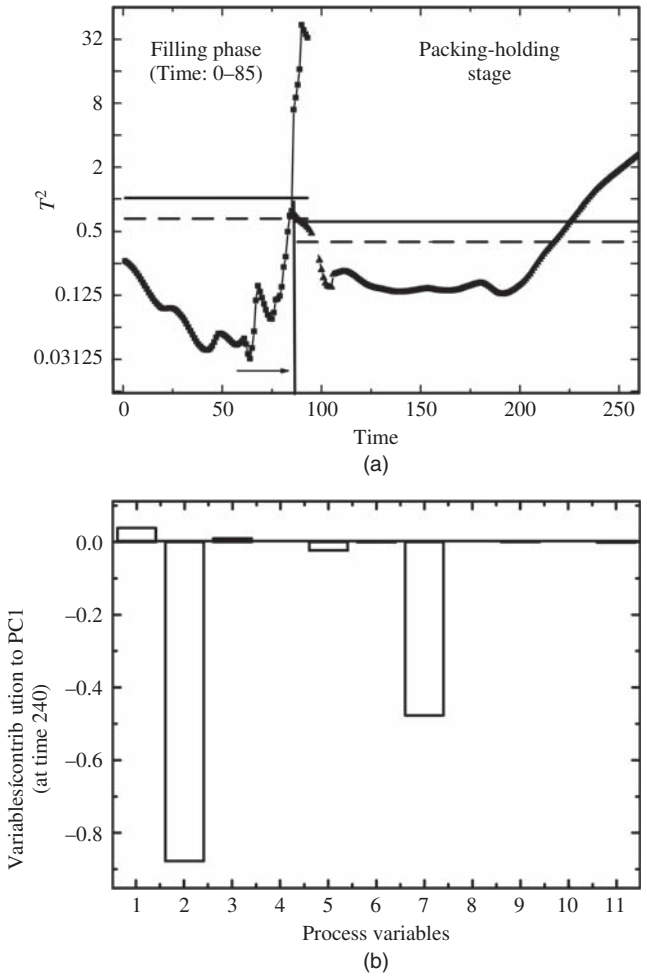

FIGURE 14.23 Monitoring and diagnosis of an abnormal batch with valve-sticking based on uneven-duration sub-PCA: (a) T^2 control chart (solid curve: 99% control limits; dash curve: 95% control limits) and (b) contribution plot.

14.4.4.3 Application of Sub-PCA with Limited Reference Data The phase-based batch process modeling and monitoring method with limited reference data has also been used to monitor the injection molding process. The variable list is shown in Table 14.3. The window length was 30 samples and the moving step was set to be 1.

The phase division and process modeling started with an arbitrary normal batch. According to the changes in local variable correlation structure along the operation time, the batch process was partitioned into several clusters, where the four major stable clusters corresponded to four steady operating phases, and other short clusters were caused by transitions between phases.

A check-ring fault that occurs in the injection phase was to be detected. As such a fault only affects the filling, packing-holding, and plastication phases, but does not influence the cooling phase, only the monitoring results of the first three phases are displayed explicitly in Figure 14.24. The control charts show that the fault was detected soon after the start of the faulty batch. When more normal batch data were available, the control limits could

TABLE 14.3 Description of the 10 Process Variables

Variable's Description	Unit
Nozzle pressure	bar
Stroke	mm
Injection velocity	mm/s
Hydraulic pressure	bar
Plastication pressure	bar
Cavity pressure	bar
Screw rotation speed	rpm
SV1 opening	%
SV2 opening	%
Mold temperature	°C

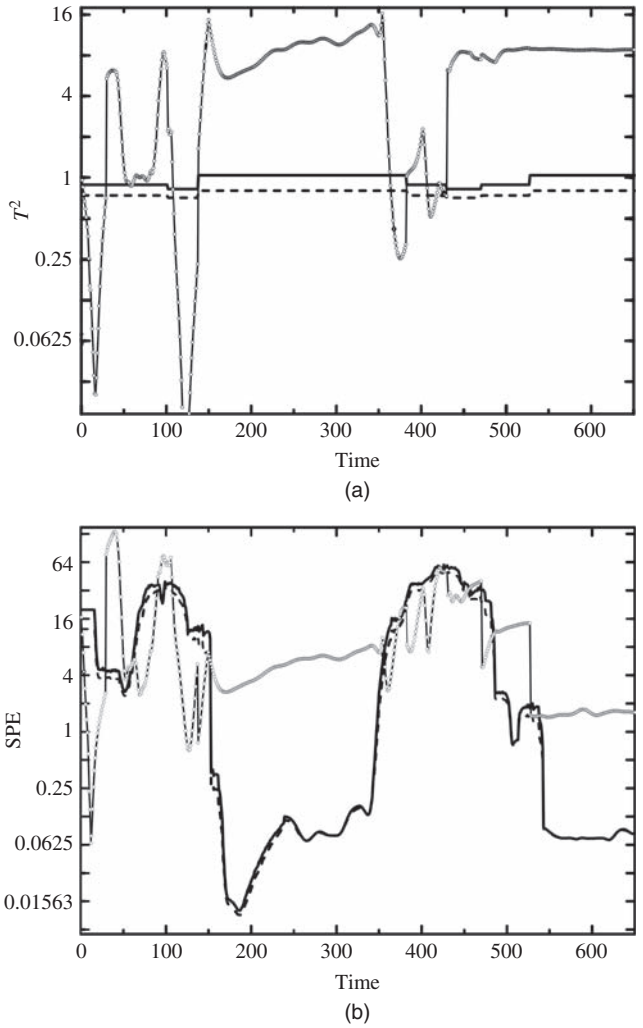

FIGURE 14.24 On-line monitoring results for a batch with check-ring problem based on sub-PCA with limited reference data: (a) T^2 control chart and (b) SPE control chart (solid curve: 99% control limits; dash curve: 95% control limits).

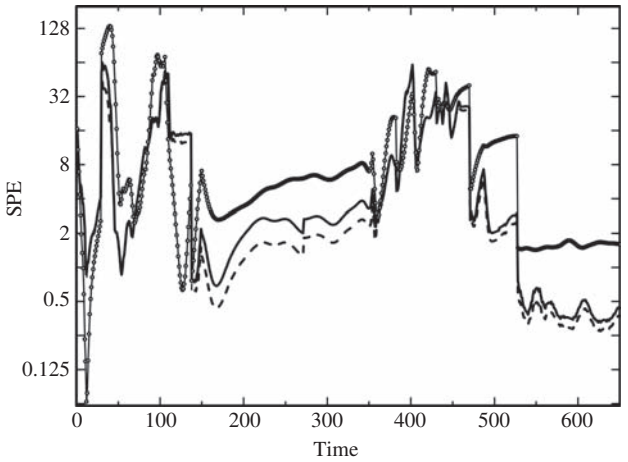

FIGURE 14.25 SPE control chart for a batch with check-ring problem based on the updated control limits (solid curve: 99% control limits; dash curve: 95% control limits).

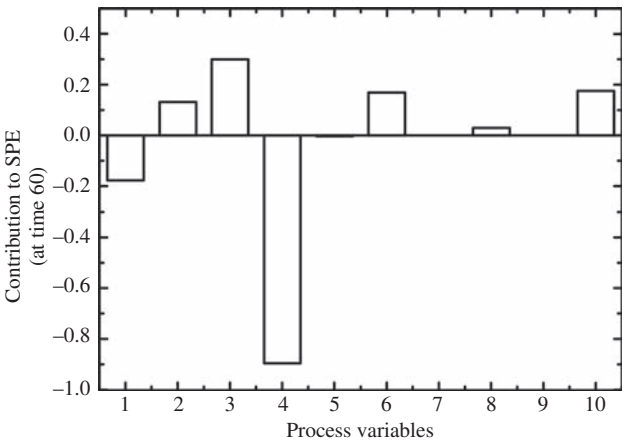

FIGURE 14.26 Contribution plot for a batch with check-ring problem based on sub-PCA with limited reference data.

be updated. The comparison between Figures 14.25 and 14.24 shows the adjustment of control limits. Contribution plot was used for fault diagnosis, as shown in Figure 14.26. The diagnosis result confirmed the engineering knowledge about the check-ring problem.

14.5 CONCLUSIONS

Although the development of APC techniques for polymer processing was quick over the last couple of years, SPC is still necessary for ensuring process safety and product quality consistency, which can be regarded as an important complementarity to APC. To deal with a large amount of variables in industries, MSPC methods have been used in the monitoring and fault diagnosis of polymer processing. The applications of the phase-based sub-PCA methods to injection molding processes are also introduced. The successful results suggest these methods can indeed enhance the process understanding, and improve fault detection efficiency and diagnosis ability.

REFERENCES

1. Venkatasubramanian V., Rengaswamy R., Yin K., et al., A review of process fault detection and diagnosis Part I: Quantitative model-based methods. Computers & Chemical Engineering, 2003. **27**(3): 293–311.
2. Shewhart W., Deming W., Statistical method from the viewpoint of quality control. 1986, New York: Dover.
3. Page E., Continuous inspection schemes. Biometrika, 1954. **41**(1–2): 100–115.

4. Page E., Cumulative sum control charts. Technometrics, 1961. **3**: 1–9.
5. Roberts S., Control chart tests based on geometric moving averages. Technometrics, 2000. **42**: 97–101.
6. Rauwendaal C., SPC - Statistical Process Control in Injection Molding and Extrusion. 2nd ed. 2008, München: Carl Hanser Verlag.
7. Jackson J., A User's Guide to Principal Components. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. 1991, New York: Wiley.
8. Jolliffe I., Principal Component Analysis. 2nd ed. Springer Series in Statistics. 2002, New York: Springer.
9. Antory D., Irwin G., Kruger U., et al., Improved process monitoring using nonlinear principal component models. International Journal of Intelligent Systems, 2008. **23**(5): 520–544.
10. Lee J., Yoo C., Choi S., et al., Nonlinear process monitoring using kernel principal component analysis. Chemical Engineering Science, 2004. **59**(1): 223–234.
11. Runger G., Multivariate statistical process control for autocorrelated processes. International Journal of Production Research, 1996. **34**(6): 1715–1724.
12. Treasure R., Kruger U., Cooper J., Dynamic multivariate statistical process control using subspace identification. Journal of Process Control, 2004. **14**(3): 279–292.
13. Wikström C., Albano C., Eriksson L., et al., Multivariate process and quality monitoring applied to an electrolysis process. Part I: Process supervision with multivariate control charts. Chemometrics and Intelligent Laboratory Systems, 1998. **42**(1–2): 221–231.
14. Choi S., Martin E., Morris A., et al., Adaptive multivariate statistical process control for monitoring time-varying processes. Industrial & Engineering Chemistry Research, 2006. **45**(9): 3108–3118.
15. Grung B., Manne R., Missing values in principal component analysis. Chemometrics and Intelligent Laboratory Systems, 1998. **42**(1–2): 125–139.
16. Imtiaz S., Shah S., Treatment of missing values in process data analysis. The Canadian Journal of Chemical Engineering, 2008. **86**(5): 838–858.
17. Nelson P., MacGregor J., Taylor P., The impact of missing measurements on PCA and PLS prediction and monitoring applications. Chemometrics and Intelligent Laboratory Systems, 2006. **80**(1): 1–12.
18. Wold S., Kettaneh N., Tjessem K., Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. Journal of Chemometrics, 1996. **10**: 463–482.
19. Westerhuis J., Kourti T., MacGregor J., Analysis of multiblock and hierarchical PCA and PLS models. Journal of Chemometrics, 1998. **12**(5): 301–321.
20. Qin S., Valle S., Piovoso M., On unifying multiblock analysis with application to decentralized process monitoring. Journal of Chemometrics, 2001. **15**(9): 715–742.
21. Chu Y., Qin S., Han C., Fault detection and operation mode identification based on pattern classification with variable selection. Industrial & Engineering Chemistry Research, 2004. **43**(7): 1701–1710.
22. Zhao S., Zhang J., Xu Y., Monitoring of processes with multiple operating modes through multiple principle component analysis models. Industrial & Engineering Chemistry Research, 2004. **43**(22): 7025–7035.
23. Kano M., Nagao K., Hasebe S., et al., Comparison of multivariate statistical process monitoring methods with applications to the Eastman challenge problem. Computers and Chemical Engineering, 2002. **26**(2): 161–174.
24. Lane S., Martin E., Morris A., et al., Application of exponentially weighted principal component analysis for the monitoring of a polymer film manufacturing process. Transactions of the Institute of Measurement & Control, 2003. **25**(1): 17–35.
25. Skagerberg B., MacGregor J., Kiparissides C., Multivariate data analysis applied to low-density polyethylene reactors. Chemometrics and Intelligent Laboratory Systems, 1992. **14**(1–3): 341–356.
26. Wold S., Cross-validatory estimation of the number of components in factor and principal components models. Technometrics, 1978. **20**: 397–405.
27. Jackson J., Mudholkar G., Control procedures for residuals associated with principal component analysis. Technometrics, 1979. **21**: 341–349.
28. Nomikos P., MacGregor J., Multivariate SPC charts for monitoring batch processes. Technometrics, 1995. **37**: 41–59.
29. Miller P., Swanson R., Heckler C., Contribution plots: A missing link in multivariate quality control. Applied Mathematics and Computer Science, 1998. **8**: 775–792.
30. Kassidas A., MacGregor J., Taylor P., Synchronization of batch trajectories using dynamic time warping. AIChE Journal, 1998. **44**(4): 864–875.
31. Yao Y., Gao F., A survey on multistage/multiphase statistical modeling methods for batch processes. Annual Reviews in Control, 2009. **33**(2): 172–183.
32. Ramaker H., van Sprang E., Westerhuis J., et al., Fault detection properties of global, local and time evolving models for batch process monitoring. Journal of Process Control, 2005. **15**(7): 799–805.
33. Nomikos P., MacGregor J., Monitoring batch processes using multiway principal component analysis. AIChE Journal, 1994. **40**(8): 1361–1375.
34. Wold S., Kettaneh N., Fridén H., et al., Modelling and diagnostics of batch processes and analogous kinetic experiments. Chemometrics and Intelligent Laboratory Systems, 1998. **44**(1–2): 331–340.
35. Rännar S., MacGregor J., Wold S., Adaptive batch monitoring using hierarchical PCA. Chemometrics and Intelligent Laboratory Systems, 1998. **41**(1): 73–81.
36. Chen J., Liu K., On-line batch process monitoring using dynamic PCA and dynamic PLS models. Chemical Engineering Science, 2002. **57**(1): 63–75.
37. Zhao C., Wang F., Jia M., Dissimilarity analysis based batch process monitoring using moving windows. AIChE Journal, 2007. **53**(5): 1267–1277.

38. Flores-Cerrillo J., MacGregor J., Multivariate monitoring of batch processes using batch-to-batch information. AIChE Journal, 2004. **50**(6): 1219–1228.
39. Garcia-Munoz S., Kourti T., MacGregor J., Model predictive monitoring for batch processes. Industrial & Engineering Chemistry Research, 2004. **43**(18): 5929–5941.
40. Lu N., Yao Y., Gao F., et al., Two-dimensional dynamic PCA for batch process monitoring. AIChE Journal, 2005. **51**(12): 3300–3304.
41. Lu N., Gao F., Wang F., Sub-PCA modeling and on-line monitoring strategy for batch processes. AIChE Journal, 2004. **50**(1): 255–259.
42. Jain A., Murty M., Flynn P., Data clustering: a review. ACM Computing Surveys, 1999. **31**(3): 264–323.
43. Ku W., Storer R., Georgakis C., Disturbance detection and isolation by dynamic principal component analysis. Chemometrics and Intelligent Laboratory Systems, 1995. **30**(1): 179–196.
44. Kourti T., Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions. Journal of Chemometrics, 2003. **17**(1): 93–109.
45. Ündey C., Ertunc S., Cinar A., Online batch/fed-batch process performance monitoring, quality prediction, and variable-contribution analysis for diagnosis. Industrial & Engineering Chemistry Research, 2003. **42**(20) :4645–4658.
46. Lu N., Gao F., Yang Y., et al., PCA-based modeling and on-line monitoring strategy for uneven-length batch processes. Industrial & Engineering Chemistry Research, 2004. **43**(13): 3343–3352.
47. Lu N., Yang Y., Wang F., et al. A stage-based monitoring method for batch processes with limited reference data, in 7th International Symposium on Dynamics and Control of Process Systems (Dycops-7). 2004, Boston.
48. Zhao C., Wang F., Gao F., et al., Adaptive monitoring method for batch processes based on phase dissimilarity updating with limited modeling data. Industrial & Engineering Chemistry Research, 2007. **46**(14): 4943–4953.
49. Kano M., Hasebe S., Hashimoto I., Statistical process monitoring based on dissimilarity of process data. AIChE Journal, 2002. **48**: 1231–1240.
50. Fukunaga K., Koontz W.L.G., Application of the Karhunen-Loeve expansion to feature selection and ordering. IEEE Transactions on Computers, 1970. **19**: 311–318.
51. Lu N., Yang Y., Gao F., et al., Stage-based multivariate statistical analysis for injection molding. Proceeding of 7th International Symposium on Advanced Control of Chemical Processes. 2004, Hong Kong. 639–644.
52. Kaspar M., Ray W., Chemometric methods for process monitoring and high-performance controller design. AIChE Journal, 1992. **38**(10): 1593–1608.